Responsiveness in Rheumatoid Arthritis. A Report from the OMERACT 11 Ultrasound Workshop


J Rheumatol 2014;41;379-382
http://www.jrheum.org/content/41/2/379

1. Sign up for our monthly e-table of contents
   http://www.jrheum.org/cgi/alerts/etoc

2. Information on Subscriptions
   http://jrheum.com/subscribe.html

3. Have us contact your library about access options
   Refer_your_library@jrheum.com

4. Information on permissions/orders of reprints
   http://jrheum.com/reprints.html

The Journal of Rheumatology is a monthly international serial edited by Earl D. Silverman featuring research articles on clinical subjects from scientists working in rheumatology and related fields.
Responsiveness in Rheumatoid Arthritis. A Report from the OMERACT 11 Ultrasound Workshop


ABSTRACT. Objective. To summarize the work performed by the Outcome Measures in Rheumatology (OMERACT) Ultrasound (US) Task Force on the validity of different US measures in rheumatoid arthritis (RA) and juvenile idiopathic arthritis (JIA) presented during the OMERACT 11 Workshop.

Methods. The Task Force is an international group aiming to iteratively improve the role of US in arthritis clinical trials. Recently a major focus of the group has been the assessment of responsiveness of a person-level US synovitis score in RA: the US Global Synovitis Score (US-GLOSS) combines synovial hypertrophy and power Doppler signal in a composite score detected at joint level. Work has also commenced examining assessment of tenosynovitis in RA and the role of US in JIA.

Results. The US-GLOSS was tested in a large RA cohort treated with biologic therapy. It showed early signs of improvement in synovitis starting at Day 7 and increasing to Month 6, and demonstrated sensitivity to change of the proposed grading. Subsequent voting questions concerning the application of the US-GLOSS were endorsed by > 80% of OMERACT delegates. A standardized US scoring system for detecting and grading severity of RA tenosynovitis and tendon damage has been developed, and acceptable reliability data were presented from a series of exercises. A preliminary consensus definition of US synovitis in pediatric arthritis has been developed and requires further testing.

Conclusion. At OMERACT 11, consensus was achieved on the application of the US-GLOSS for evaluating synovitis in RA; and work continues on development of RA tenosynovitis scales as well as in JIA synovitis. (First Release Nov 15 2013; J Rheumatol 2014;41:379–82; doi:10.3899/jrheum.131084)

Key Indexing Terms:
ULTRASOUND RESPONSIVENESS RHEUMATOID ARTHRITIS OMERACT
The OMERACT Ultrasound Task Force met for the first time at OMERACT 7 in 2004 with the aim of investigating the role of ultrasound (US) in rheumatology, and developing its metric properties, particularly in the field of inflammatory arthritis. The first step was to develop and publish standardized definitions of US pathologies. The group then performed a systematic review of the metric properties of US for the detection of synovitis at the individual joint level (for multiple peripheral joints) that highlighted gaps in the literature, including a lack of reliability data with respect to intraobserver, interreader, and interobserver reliability. Following this work, a series of iterative intraobserver and interobserver reliability exercises, the group developed and tested definitions and scoring of joint pathologies related either to inflammation or to structural damage in rheumatoid arthritis (RA). This subsequently resulted in the ongoing development of a US GLObal Synovitis Score (US-GLOSS), at the whole patient level, that would be proposed for application in multicenter studies applying either extensive or reduced US joint counts have shown the feasibility of US for following patients under treatment. However, many differences have been reported in the scoring system used (i.e., greyscale synovitis, Doppler, both modalities) as well as in the number of joints assessed and in the correlations (construct validity) between US scores and clinical and laboratory findings. Thus, development of a US-GLOSS required testing of sensitivity to change of the OMERACT synovitis scoring system in a multicenter study.

During the OMERACT 11 Workshop, results were presented of the application of the US-GLOSS in patients with RA. The study represented the first international, multicenter, prospective, open-label therapeutic trial that used the OMERACT US-GLOSS and tested the response to treatment with abatacept and methotrexate (MTX) in patients with RA and inadequate response to MTX (manuscript in preparation). In this study, US responsiveness was used as the primary objective. US-GLOSS, by combining synovial hypertrophy and power Doppler (PD) signal in a composite score applied at joint level, was able to demonstrate, at patient level, early signs of improvement in synovitis from days 7, 15, and 29, increasing up to Month 6. A reduced subset of joints (GLOSS) that best represented the global PDUS score for paired 22 joints of all patients with RA demonstrated positive results at the joint level. In addition, recent longitudinal studies applying either extensive or reduced US joint counts have shown the feasibility of US for following patients under treatment. However, many differences have been reported in the scoring system used (i.e., greyscale synovitis, Doppler, both modalities) as well as in the number of joints assessed and in the correlations (construct validity) between US scores and clinical and laboratory findings. Thus, development of a US-GLOSS required testing of sensitivity to change of the OMERACT synovitis scoring system in a multicenter study.

The concept of developing a US-based scoring system for synovitis in RA at the patient level was agreed on through group discussions and feedback sessions at OMERACT 10 in 2010. This decision was based on the need for integrating the components of synovitis (i.e., hypoechoic synovial hypertrophy and synovial Doppler signal) in a unique global score of joint inflammatory activity in RA. This scoring system, applied at the patient level by means of a multit joint US assessment, should allow an objective followup of patients under treatment and should provide a feasible and objective instrument for evaluating disease activity in patients with RA. The work developed by the group in the assessment of reliability in the detection and scoring of RA synovitis demonstrated positive results at the joint level. In addition, recent longitudinal studies applying either extensive or reduced US joint counts have shown the feasibility of US for following patients under treatment. However, many differences have been reported in the scoring system used (i.e., greyscale synovitis, Doppler, both modalities) as well as in the number of joints assessed and in the correlations (construct validity) between US scores and clinical and laboratory findings. Thus, development of a US-GLOSS required testing of sensitivity to change of the OMERACT synovitis scoring system in a multicenter study.

Following the decision taken during the OMERACT 10 meeting, the group expanded the validation of US in RA by working on tenosynovitis. Thus, starting from a systematic literature review of US definitions, scoring systems, and validity for tendon involvement in RA, a series of reliability exercises was performed that focused on the detection and scoring of tendosynovitis and tendon damage in patients with RA. It was also agreed at OMERACT 10 to work on US as an outcome measure in pediatric arthritis. Based on a systematic literature review, the group started work on the definition of a normal joint that will lead to definitions of pediatric joint synovitis. At the same time future areas of standardization were agreed on.

The aim of the present article is to summarize the recent work performed by the OMERACT US Group in the assessment of responsiveness for a number of measures in RA, as presented during the OMERACT 11 Workshop.
questions for US synovitis were presented to participants for potential endorsement. The first voting question: “Do you support continuing to evaluate US in assessing synovitis by using greyscale and Doppler modalities separately?” was endorsed by > 80% of the OMERACT delegates to have appropriate validation data. The second question: “In addition, do you support to continue development and subsequent validation of a combined score?” received 70% of the vote. The third question: “Do you support the idea to improve the feasibility of US assessment by developing and validating a reduced joint count?” was endorsed by > 80% of delegates. Finally, the fourth question: “Do you support implementation of US as a secondary/exploratory outcome in therapeutic clinical trials?” received 80% of the vote.

Tenosynovitis
The work on tenosynovitis was essentially focused on the assessment of discrimination of US for detecting and scoring tendon involvement in RA. The starting point was a systematic literature review of US definitions, scoring systems, and metric properties for tenosynovitis and tendon damage in RA4. Then, based on the limited available data on reliability, the group started a series of exercises focused initially on testing the reliability of US in detecting lesions, and subsequently aimed at assessing the reliability of US in scoring tenosynovitis and tendon damage5,6. A Delphi process was previously performed on defining the scoring systems for tenosynovitis6. During the US Workshop at OMERACT 11, results of the literature review were presented with focus on the different aspects of validity4. The results of the intraobserver and interobserver reliability exercises were also reported: moderate-to-good intraobserver and interobserver reliability for tenosynovitis and tendon damage were demonstrated for both greyscale and PD. Then, following the above Delphi consensus process (which examined agreement of a group of experts on US-defined tenosynovitis and US scoring systems of tenosynovitis in RA), a second patient-based reliability exercise was performed. The results showed good intraobserver reliability and moderate to good interobserver reliability for B-mode and PD modalities5. A formal research agenda was also developed including a third patient-based reliability exercise on tendon damage in RA, performed in June 2012 in Amsterdam, with the aim of developing an OMERACT atlas on scoring tenosynovitis and tendon damage in RA.

Pediatric Arthritis
Several publications have highlighted the capability of US to detect joint involvement in pediatric arthritis revealing the high frequency of asymptomatic findings. At OMERACT 11 the group reported results of a systematic literature review demonstrating that validity of US in pediatric arthritis had not yet been established and standardized7. Using results of an international survey on the use of US in pediatric rheumatology as the basis for future work, the first identified area for standardization was the definition and detection of US synovitis in juvenile idiopathic arthritis. Such pathology does not differ from that observed in adults; however, the appearance of the joint can vary according to the age and maturation of the child, with the presence of variable amounts of vascularized epiphyseal cartilage. This makes the distinction of normal findings from synovitis more challenging. The work of the group has also been focused on developing definitions of normal joint components for different age groups through a Delphi consensus process and by testing them in a reliability exercise involving healthy children, performed in Madrid in March, 2012. The preliminary results were presented during the session. The group is now developing preliminary definitions for common US-detected pathologies in pediatric rheumatology diseases.

At OMERACT 11 all aspects presented on synovitis were strongly endorsed by OMERACT delegates, further supporting US as a good outcome measure for evaluating activity and responsiveness in RA. Consensus was achieved on the application of a US-GLOSS scoring system for synovitis in RA and its use in therapeutic clinical trials. Further randomized controlled studies are needed to confirm these results and explore predictive validity in terms of structural damage and stratification of “patients at risk” for severe disease.

REFERENCES
8. Naredo E, Wakefield R, Iagnocco A, Terslev L, Filippucci E,